DOCUMENT RESUME

ED 170 375 TM 008 971

AUTHOR Cooper, James M.; Jones, Howard L.
TITLE Formative Evaluation Efforts in a Teacher Education Program.
PUB DATE Apr 79
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Curriculum Evaluation; *Evaluation Criteria; *Evaluation Methods; Evaluation Needs; *Formative Evaluation; Higher Education; Participant Satisfaction; *Performance Based Teacher Education; Program Evaluation; *Teacher Education Curriculum
IDENTIFIERS University of Houston TX

ABSTRACT

An undergraduate competency-based teacher education curriculum was evaluated after it had been in effect for three years. To generate a workable system for the collection and use of formative evaluation data, a list of program characteristics were identified, as well as a set of teaching skills and competencies relating to subject knowledge. Two approaches were used to identify how well these elements had been implemented: discrepancy evaluation between current practices and program goals, and identification of the satisfaction and successes of students and graduates. Twelve characteristics of the successful teacher education system were identified as program goals, as well as indicators that these characteristics had been implemented. Data were collected from questionnaires administered to students, instructors, supervisors, and supervising teachers, and from the analysis of relevant documents--descriptions of instructional modules, grading policies, and program policy statements. This information was disseminated in order to encourage curricula improvement. The success and satisfaction of students and graduates were studied from the results of high-inference rating scales administered to student teachers and their supervisors. (Results are briefly described). (GDC)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED170375

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# FORMATIVE EVALUATION EFFORTS IN A TEACHER EDUCATION PROGRAM

James M. Cooper and Howard L. Jones

University of Houston

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James M. Cooper
Howard L. Jones

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

A paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April, 1979

IM008 971



Session 25.11

Need for Formative Evaluation

Teacher education program evaluation stands at an interesting and critical point today. It is generally agreed that past evaluation efforts have been minimal with respect to both quantity and quality. For many institutions program evaluation means sending out a questionnaire to a sample of program graduates shortly before the next NCATE visitation. While this practice perhaps met the letter of the law with respect to the old NCATE standards, the new 1979 NCATE Standard 6 requires much more in the way of program evaluation. Specifically, the standard states:

> Maintenance of acceptable teacher education programs demands a continuous process of evaluation of the graduates of existing programs, modification of existing programs, and long-range planning. The faculty and administrators in teacher education evaluate the results of their programs, not only through the assessment of graduates but also by seeking reactions from persons involved with the certification, employment, and supervision of its graduates. The findings of such evaluation are used in program modification.

Program evaluation is meant to be an ongoing process, one that aids decision makers in refining their programs. Its greatest power lies in its formative uses.

TenBrink defines evaluation as "...the process of obtaining information and using it to form judgments which in turn are to be used in decision making."[1] The three key elements are information, judgments, and decisions, with the kinds of judgments and decisions needing to be made dictating the information needs. The developers of each program must decide what questions they want to answer before collecting any information. In this

[1] Terry D. TenBrink, Evaluation: A Practical Guide for Teachers, New York: McGraw-Hill Book Co., 1974, p. 8.

way, they ensure that the data they do collect is relevant to their needs; decisions can thus be more validly data-based.

Potential Roadblocks/Problems to Effective Formative Evaluation

Faculty and staff of most institutions of higher education would generally agree that their institutions need more formalized means for collecting formative data about their programs. However, most institutions have loosely organized efforts to collect minimal data. Course and faculty evaluations by students are generally completed after each semester or quarter. In some programs there is a concerted effort to gather summative assessment data at the completion of internship or student teaching.

These same faculty and staff generally would agree that the student and faculty assessments of "during-program" stages is unnecessarily limiting. In a recent survey, it was noted that 80+% of teacher education stitutions valued the gathering of data about their teacher graduates. In the same survey, however, it was noted that 49+% of the institutions were not gathering any follow-up data.[2]

The position presented in this paper is that effective formative evaluation must be made at all levels of the program - before the students start, during the program, and after the graduates start their professional careers. Part of this position is the acceptance of the fact that there are roadblocks and problems that must be solved by any institution before effective procedures can be developed and implemented. Among these problems are the following:

[2] Personal communication from Gary Peer, Assistant Dean of Education, University of Tulsa, to Edell Hearn, Dean, College of Education, Tennessee Technological University, February 28, 1978.

1. Overcoming Inertia and Reluctance

Most faculty, staff and students of teacher education programs have not been engaged in the gathering of formative data about their efforts. How can students be convinced that their suggestions, complaints and even satisfactions will be used by program developers in a useful manner? How can students be convinced that their ideas will be taken seriously? How can faculty members be convinced that formative data is more than "fuel for paranoia?" How can faculty members be helped to deal with the opportunity to improve their offerings rather than looking for another course that is "worse", thereby evidencing the "I'm OK compared to him" syndrome? How can graduates be convinced that it is important to take the extra effort to provide feedback to their alma mater for the purpose of changing elements?

2. What Kinds of Data Should be Gathered?

What data are most useful for the program? Which data are the most timely, persuasive and relevant? General "satisfaction" data are usually diffuse and difficult to use to change programs. Rating scales or similar satisfaction scales really do not permit a "fine tuning" of program offerings, yet these data are the easiest to gather. The gathering of personality characteristics data is usually a stressful effort, yet these data might provide some indication of program effects on students. Are these data worth the psychological and other costs necessary to gather them? Is it important to gather low inference data about the classroom mannerisms of student teachers or graduates? Finally, how much data is enough? Program developers and evaluators must

recognize that over-gathering data can have a deleterious effect on students' satisfactions.

3. Who Will Make the Key Decisions?

Teacher education programs at most institutions are loosely organized, diffuse structures. Accountability and responsibility for individual progress, not to mention the progress of entire programs, are often very nebulous. A key element in formative evaluation is the identification of who is in charge of the evaluation. Who will make the decisions on the selection/development of assessments? Who will administer instruments or collect data within teachers' classrooms? Who will help program builders and implementers interpret the data that are gathered? In short, who owns what decisions and who will have both the responsibility and the accountability of the evaluation effort?

4. What Strategies Will Be Used to Gather the Data?

The successful gathering of useful data takes planning; resources (fiscal and personal) are usually scarce or unavailable. Where these resources come from and the assurance that there is maximum synergy in the collection of data will depend on the viability of planned strategies. Where will the finances come from? Will faculty be provided released time to gather data, especially data from the field? How can assurances be made that no one set of students or faculty is overburdened by data gathering efforts?

These are but a few of the problems and issues that must be considered if formative evaluation of teacher education programs is to work successfully.

Formative Evaluation Efforts at the University of Houston

In recognizing the importance of continuous formative evaluation in teacher education, a task force of faculty members was formed in 1977 to develop and pilot test an evaluation system for the undergraduate offerings of the College of Education at the University of Houston. 1977 was an especially critical year for the College. In 1971 the first experimental CBTE program was launched, serving about 10% of the undergraduate students in the College. In 1974, CBTE was ratified as the one instructional system for the entire college. Thus, three years later (1977) it was considered time to generate a workable system for the collection and use of formative data.

The developed plan for evaluation was an expansion of past program assessment procedures at the University of Houston. Throughout its history of teacher education there have been attempts to determine the satisfaction of teacher education students with their program as well as the success of graduates as teachers. However, recent changes in programs made it imperative to consolidate the informal evaluation processes. Between 1974-77, faculty and staff of the College of Education agreed to a set of programmatic assumptions to guide the development and implementation of their teacher education program. In addition, a list of program characteristics, as well as a set of generic competencies relating to special subject matter knowledge and teaching skills, were identified.

While the assumptions, program characteristics and competencies noted above were agreed to in principle, the level of implementation of these elements was unknown. The purpose of this research effort was to identify how well these elements had been implemented.

Specifically, the initial program evaluation effort had two major goals; each identified by an evaluation sub-system.

Sub-system 1: A discrepancy evaluation between what presently occurs in the undergraduate teacher education program and those programmatic characteristics that should be evident given the agreed upon assumptions about what constitutes an effective instructional system.

Sub-system 2: The identification of the satisfaction and successes of both present and former teacher education students at the University of Houston.

## SUB-SYSTEM 1 - PROGRAM CHARACTERISTICS

In 1975 the faculty of the College of Education formally adopted twelve assumptions regarding instructional system characteristics believed to be essential to the operation of an effective teacher education program. These twelve characteristics formed the foundation upon which the total instructional system was to be based. This section of the paper describes an evaluation effort to determine the extent to which the twelve instructional characteristics which had been adopted in 1975.

The twelve characteristics believed essential to an effective teacher education program are that the program be:

1. Competency based; that is, the competencies to be demonstrated by the student are made explicit, the criteria to be applied in assessing the student's competencies are made explicit, and the student is held accountable for meeting those criteria. The emphasis is on the demonstration, not the acquisition, of those competencies specified as program expectations.

2. Campus-centered and field-oriented; that is, students are provided with opportunities to experience instruction and to demonstrate competence both on campus and in field sites depending upon the nature of the particular instructional or assessment activity, and a determination as to where that activity can most effectively and efficiently take place.

3. Role-model based; that is, a conceptual model of the teacher's role is used as a basis for identifying and specifying those competencies students are expected to demonstrate.

[3] This evaluation effort is reported in more detail in: Wilford A. Weber and James M. Cooper, "Evaluation of Instructional System Characteristics of the Professional Teacher Preparation Program at the University of Houston." in Teacher Education Program Evaluation and Follow-Up Studies: A Collection of Current Efforts, Shirley M. Hord and Gene E. Hall (eds.). Austin: University of Texas R & D Center for Teacher Education, 1978, pp. 49-66.

4. Criterion-referenced; that is, assessment procedures are designed to determine whether or not a student has demonstrated competency at or above the level of mastery specified; the competency of each student is judged on the basis of predetermined criteria and is not determined through a comparison involving other students.

5. Pluralistic; that is, there is conscious acceptance of the notion that no one philosophy of instruction has been proved best; consequently, a variety of divergent views are posited as untested assumptions which the system must test.

6. Humanized; that is, program design and operation recognize the dignity and worth of each individual so that each perceives that he or she is being treated as one of worth.

7. Personalized; that is, the student is provided with instruction which takes into account his or her uniqueness.

8. Modularized; that is, the delivery of instruction is accomplished through the utilization of instructional modules; instructional modules are sets of learning activities--rationale, objectives, prerequisites, pre-assessment procedures, learning alternatives, post-assessment procedures, and remedial procedures--which are intended to facilitate the learner's acquisition and demonstration of a particular competency or set of competencies.

9. Multi-institutional; that is, the design and operation of the teacher education program is a responsibility shared by colleges, public schools, and the organized teaching profession.

10. Systemized; that is, the systems approach is used in program design and operation.

11. Regenerative; that is, the program is an open system capable of continuous revision on the basis of constructive data supplied by sound formative and summative evaluation procedures.

12. A single system with alternatives within it; that is, the program is a single instructional system which constitutes an integrated, comprehensive whole; however, the system does accommodate a divergence of viewpoints within its confines.

Data Needs and Collection Procedures

A sub-group of the evaluation task force generated sets of indicators which one might expect to find if each of the twelve instructional system characteristics had been operationalized. Those indicators which appeared to be most important and for which it appeared feasible to collect data during the spring 1977 semester were agreed upon by the task force.

The indicators about which data were to be collected were of two types: (1) perception; and (2) documentation. It was determined that perceptual data concerning the twelve characteristics should be collected from students, instructors, supervisors, and supervising teachers associated with the program. Several types of data-gathering procedures were considered to be appropriate for this purpose. However, it was decided that these data, given feasibility considerations, would best be collected through the use of questionnaires--a reasonably good means of collecting perceptual data. This decision led to the construction of four questionnaires, one for each of the four groups to be sampled. Additionally, it was determined that data concerning the characteristics might be gained through analyses of various program documents, policy statements, and instructional materials. It was decided that the use of checklists might be an effective means for this purpose since a checklist can be designed to provide fairly easily obtained, objective information as to whether or not a given characteristic is present.

Questionnaires were mailed to all supervising teachers and student teachers toward the end of the spring semester. At about the same time, questionnaires for the instructors and university supervisors were distributed by instructors in selected classes so that a copy might reach every student enrolled in the PTPP. Questionnaires were returned by way of the same channels.

In order to obtain the program documents which were to be analyzed, the sub-group provided the associate dean for undergraduate studies with a list of each of those documents which was needed. The associate dean, in turn, requested various individuals with Professional Teacher Preparation Program responsibilities to provide the specified documentation. These documents were to be sent to the associate dean, who passed them on to the sub-group for analysis.

Some Limitations

Before briefly describing a few of the study's findings, it is perhaps wise to consider several of the factors which may limit those findings. First, the findings reported here are valid to the extent that the indicators chosen to represent the program characteristics do in fact reflect those characteristics. Second, the accuracy of the findings is influenced by the extent to which the data chosen for analysis represent the selected indicators. Third, the findings and conclusions are limited by the fact that the return rate on the questionnaires and the requested documentation was only moderately high. Fourth, the findings are limited by the reliability and validity of both the questionnaires and the analysis of documents; no attempt was made to prepare statistical estimates of their validity and reliability. And fifth, the findings are limited to the extent that questionnaire responses and program documentation were accurate and honest.

Space precludes reporting more than just a few of the findings from the questionnaire data and the document analyses. Several of the more interesting items are described here to provide examples of the kinds of data obtained and the manner in which they were treated.

Questionnaire Data. The first twelve items on each of the four questionnaires which were administered to the students, instructors, supervisors, and supervising teachers in this study consisted of statements which presented each of the

program's assumptions concerning an effective instructional system. Using a nine-point scale, with one being strongly disagree and nine being strongly agree, respondents were asked to indicate the degree to which they felt that the program had been consistent with each assumption. For example, the first statement on each of the four questionnaires was concerned about the extent to which respondents perceived the program to be competency based.

The statement to which respondents were to react said: "The program is competency based; that is, the competencies students are asked to demonstrate are made explicit, the criteria applied in assessing the competencies of students are made explicit, and students are held accountable for meeting those criteria.

Table 1 presents the results obtained from students, instructors, supervisors, and supervising teachers for those first twelve items. Mean score data suggest that all four groups agreed that the program had operated in a manner which was consistent with the stated assumptions. Mean scores ranged from a low of 5.13 to a high of 7.85. Student mean scores ranged from 5.99 (personalized) to 7.29 (campus-centered and field-oriented) with the other ten scores falling between 6.10 and 7.04, a rather small range.

Instructor mean scores ranged from 5.13 (personalized) to 7.64 (campus/field); overall they were a bit lower than student mean scores with five scores ranging from 5.13 to 5.96, and another six scores ranging from 6.02 to 6.71. Supervisor mean scores ranged from 5.38 (role-model based) to 7.47 (humanized); four of these ranged from 5.38 to 5.78, while seven ranged from 6.23 to 6.96. Overall, the mean scores obtained from supervising teachers were the highest; they ranged from 6.33 to 7.85.

These results were encouraging. While the data suggest that there was much room for improvement, they also indicated progress. The data from university personnel--instructors and supervisor, those most closely involved in the program's development--indicated that they agreed the program was operationalized in a way

Table 1

Summary of Student, Instructor, Supervisor, and Supervising Teacher Questionnaire Data for All Twelve Program Characteristics, Variables 1 through 12. Items 1 through 12

| Program Characteristic | Students | | | Instructors | | | Supervisors | | | Supervising Teachers | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| Competency Based | 510 | 6.55 | 1.99 | 44 | 6.55 | 1.66 | 23 | 5.61 | 1.88 | 213 | 6.33 | 2.31 |
| Campus/Field | 503 | 7.29 | 1.70 | 44 | 7.64 | 1.66 | 23 | 6.65 | 2.08 | 197 | 7.26 | 1.88 |
| Role-Model Based | 493 | 6.24 | 1.96 | 42 | 5.88 | 1.99 | 21 | 5.38 | 2.22 | 197 | 6.80 | 1.94 |
| Criterion-Referenced | 496 | 6.49 | 2.11 | 43 | 6.02 | 2.45 | 23 | 6.96 | 2.16 | 206 | 6.77 | 2.20 |
| Pluralistic | 503 | 7.04 | 2.10 | 43 | 6.05 | 2.25 | 23 | 6.91 | 2.15 | 206 | 7.42 | 1.92 |
| Humanized | 509 | 6.92 | 2.04 | 44 | 6.41 | 2.43 | 23 | 7.47 | 1.90 | 212 | 7.85 | 1.72 |
| Personalized | 507 | 5.99 | 2.37 | 45 | 5.13 | 2.70 | 23 | 5.78 | 2.65 | 196 | 6.95 | 2.21 |
| Modularized | 501 | 6.98 | 1.95 | 45 | 6.71 | 1.97 | 22 | 6.23 | 2.14 | 190 | 6.52 | 2.14 |
| Multi-Institutional | 510 | 6.98 | 1.80 | 45 | 5.96 | 2.54 | 23 | 6.39 | 2.54 | 212 | 7.73 | 1.79 |
| Systemized | 508 | 6.33 | 2.05 | 45 | 6.11 | 2.35 | 23 | 6.65 | 2.04 | 197 | 7.17 | 1.83 |
| Regenerative | 491 | 6.35 | 2.09 | 45 | 5.64 | 2.59 | 22 | 6.77 | 2.07 | 191 | 7.35 | 1.77 |
| Single System | 505 | 6.10 | 2.03 | 45 | 5.24 | 2.52 | 22 | 5.73 | 2.12 | 193 | 6.90 | 1.99 |

which was consistent with the specified assumptions. However, they were the groups which perceived the greatest lack of congruence between program intent and program reality. They seem to have said that "we're getting there but we're not there yet." Student data suggest that students also agreed that the program was reasonably consistent with the assumptions on which it was based. However, the data also suggest areas needing improvement. On the other hand, data from the supervising teachers seem to indicate that they felt the program was well on its way to "getting where it wanted to go." One might speculate that their views are colored by their experiences with other teacher education programs, including those in which they were undergraduates.

In summary, then, the data from the first twelve items on each of the four questionnaires, as presented in Table 1, suggest that while progress had been made in operationalizing the instructional system characteristics there existed much room for improvement.

In an attempt to gain a greater understanding of student perceptions, student questionnaire data were analyzed according to the respondent's certification program (elementary education; secondary education; art education; health and physical education; or music education), and his or her standing (Phase I, II, III, or IV). This type of analysis made it possible to see if sub-groups of students in the program perceived it similarly or differently.

The first twelve items of the questionnaire provided data relevant to broad areas, while the other items focused on specific aspects of the program, particularly the methods courses, foundations of education course, introductory courses, and the multicultural course. For example, the data reported in Table 2 indicate that the foundations of education course (FED 361) was perceived by students to have been considerably less competency based than the other three. Similarly, data reported in Table 3 indicate that the foundations of education

Table 2

Summary Data from Student Questionnaire for Characteristic 1
(Competency Based)
Items 36 through 51

| Item | Course | N | Mean | SD |
|---|---|---|---|---|
| 36 | FED 361 | 391 | 5.27 | 2.87 |
| 40 | FED 361 | 393 | 5.15 | 2.85 |
| 44 | FED 361 | 386 | 5.31 | 2.85 |
| 48 | FED 361 | 385 | 5.77 | 2.80 |
| 37 | C&I 362-333 | 476 | 7.20 | 2.06 |
| 41 | C&I 362-333 | 480 | 6.00 | 2.08 |
| 45 | C&I 362-333 | 480 | 6.86 | 2.20 |
| 49 | C&I 362-333 | 462 | 7.08 | 2.11 |
| 38 | C&I 335 | 210 | 7.04 | 2.05 |
| 42 | C&I 335 | 205 | 6.71 | 2.31 |
| 46 | C&I 335 | 208 | 6.75 | 2.32 |
| 50 | C&I 335 | 201 | 6.90 | 2.21 |
| 39 | Methods | 342 | 6.94 | 2.09 |
| 43 | Methods | 337 | 6.77 | 2.09 |
| 47 | Methods | 330 | 6.84 | 2.01 |
| 51 | Methods | 323 | 6.88 | 2.04 |

course was also perceived to be considerably less modularized than the other three groups being rated. These findings generated great concern regarding the structure of the foundations of education course.

Table 3

Summary Data From Student Questionnaire for Characteristic 8
(Modularized)
Items 114 through 117

| Item | Course | N | Mean | SD |
|---|---|---|---|---|
| 114 | FED 361 | 360 | 3.57 | 2.92 |
| 115 | C&I 362-333 | 454 | 7.09 | 2.20 |
| 116 | C&I 335 | 194 | 7.17 | 2.05 |
| 117 | Methods | 310 | 6.22 | 2.57 |

Document Analyses

In addition to the collection and analysis of perceptual data, the evaluation--as noted earlier--included an analysis of various pertinent program documents. As with the questionnaire data, the intent was to look for evidence that the program characteristics which had been specified as assumptions had been operationalized. The sorts of documents which were requested included:

1. Descriptions of all instructional modules used in PTPP. These were examined for evidence of competency based instruction, modularized instruction, and criterion-referenced assessment, for example.

2. Descriptions of faculty, program area, and department grading policies and copies of all course grades given to students during the fall 1976 semester. These were examined for evidence of criterion-referenced assessment procedures and grading, for example.

3. Program policy statements. These contained descriptions of formally adopted goals, assumptions, and required student competencies. These were examined for evidence of a number of the characteristics.

The major findings from the document analyses was that less than half of the PTPP courses were modularized and of those which were, the modules were often deficient in terms of the components they were to contain. Table 4 presents the data for C&I 362: Introduction to the Profession of Teaching, a course for which the findings were rather typical.

Data Utilization

Findings from the module descriptions analyses were made available to program managers in June 1977. They used the findings to begin to identify priority areas needing improvement. As a result, a task force was formed in the summer of 1977 and charged with responsibility for revising instructional modules for C&I 362 and 333. The products of that task force, thirty-nine modules, were designed so as to ensure that each contained all of the components instructional

Table 4

An Analysis of the Instructional Modules: C&I 362. Introduction to the Profession of Teaching

| Module | Rationale | Objectives | Prerequisites | Pre-Assessment | Learning Alternatives | Post-Assessment | Remediation |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 2 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 3 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 4 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 5 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 6 | 1 | 2 | 2 | 2 | 2 | 2 | 0 |
| 7 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 8 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Code: 0 = Component not present.
1 = Component present; did not meet criteria.
2 = Component present; met criteria.

modules were expected to contain, and to ensure that all components met the criteria which had been established.

A summary report of the evaluation findings was made available for distribution to all faculty involved in the PTPP. Additionally, program managers held individual conferences with faculty members in various program areas to share data relevant to those program areas.

## SUB-SYSTEM-2 SUCCESS AND SATISFACTION OF STUDENTS

Program developers face a dilemma in the determination of the success of program graduates. Program graduates, both those completing student teaching and in-service teachers, should have a positive impact on their pupils. By assessing this impact using standardized or other tests, an indicator of teacher success would be evident. This "accountability" model is considerably easier to discuss than it is to implement, however. The variance of school settings, class sizes, available teaching aids, pupil aptitudes and motivations all deter true accountability evaluations.

A second possibility of evaluating PTPP products - student teachers and in-service teachers - is that of collecting data on the teachers as they teach. This step would call for classroom observations and the use of high or low inference instruments to evaluate the teachers' interactions with pupils.

Difficulties with the implementation of this effort are also evident. The time needed for actual classroom data collection is only one difficulty. Furthermore, the UH PTPP program has placed a high positive value on the use of clinical supervisory modes. In such modes, supervisors work with student teachers and School-Based Teacher Educators (cooperating teachers) in assessing, in a formative mode, the processes of the student teachers. By implementing what could be perceived as a "big brother" approach to data collection, it was felt that much of the positive affect associated with the clinical supervision mode would be damaged.

To collect data on the success of students it was decided that a high-inference rating scale would be used. The scale would be administered to student teachers and their supervisors. In addition, the School Based Teacher Educator in whose classroom the student teacher worked would be asked to rate the student teacher on the scale items.

In April, 1977, near the completion of the student teaching experience, all student teachers, their supervisors and School Based Teacher Educators were administered a rating scale to assess the success of the student teacher.

The instrument consisted of a series of 1-5 Likert scales on which the rater identified his/her ratings. The items reflected 15 of the 18 generic[4] teaching competencies of the undergraduate program - competencies that all students in all programs are expected to master. Because a few competencies are quite broad, these competencies were assessed in two or more items. For example, Competency Eight (Promotes Effective Patterns of Communication), was subdivided into four items - Communicates with Students; Communicates with Other Teachers; Communicates with Supervisors; Communicates with Parents).

Furthermore, to identify some level of perceived importance of each of the competencies, raters were asked to identify (again on a 1-5 scale) the importance that they associated with the statement. In cases where raters felt they could not respond to an item they were instructed to indicate "Can't Rate".

In essence, raters (students, supervisors, School Based Teacher Educators) indicated their perception of student teacher's demonstration of program competencies, as well as identifying their personal feelings about the value of each

[4]Three additional competencies have been added to the curriculum of the PTPP program: (16) incorporation of career education concepts into instruction: (17) knowledge and skills pertaining to mainstreaming; (18) demonstration of adequate content knowledge.

competency statement. Anonymous responses were received from 191 student teachers, 22 University supervisors of the student teachers, and 223 School Based Teacher Educators.

PTPP Success Study - Student Teachers

Table 5 reports statistical data for each of the Perceived Importance and Student Teacher Success parts of the PTPP Questionnaire. In viewing the entire set of data several elements are evident:

1. There appears to be general agreement with the importance of program competencies. Not one competency item was rated below 4.124 by any of the three rating groups. Thus, there appears to be a general consensus of programmatic thrusts. This should not be surprising considering the effort spent in acquiring SBTE and faculty input into the program. However, it is rewarding to see the agreement.
2. In all cases of the rating of student performance, student teachers generally rated themselves higher than their corresponding SBTEs and University supervisors. It is difficult to account for this since students were assured anonymity and were told that their reports would not count on grading, etc.
3. In viewing each of the tables, it appears that Elementary Education SBTEs and student teachers were consistently rating the competencies and student performance higher than the University supervisor. Again, it is difficult to account for this.

Because of the difficulties in obtaining truly reliable ratings across groups on student performance and perceived importance of the competency statement, an additional step was taken. The mean ratings for each group on each of the scales were ranked within group. For example, SBTE mean ratings of all

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | tem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 15.0 | 4.488<br>18 | 4.727<br>13(T)* | 4.611<br>14(T)* | 1. Apply knowledge of human growth and development in planning for instruction. | 4.167<br>21 | 3.727<br>20 | 4.033<br>18 | 19.7 |
| 9.7 | 4.699<br>10 | 4.909<br>6(T)* | 4.617<br>13 | 2. In planning for instruction apply knowledge about how students learn. | 4.154<br>22 | 3.682<br>21 | 4.037<br>17 | 20 |
| 20.3 | 4.326<br>20 | 4.409<br>19 | 4.140<br>22 | 3. Write clearly stated lesson and unit objectives. | 4.410<br>12 | 4.046<br>11 | 4.192<br>11 | 11.3 |
| 16.7 | 4.571<br>13 | 4.454<br>18 | 4.426<br>19 | 4 Construct understandable and useful lesson and unit plans. | 4.484<br>8 | 4.182<br>9 | 4.252<br>7 | 8 |
| 19 | 4.535<br>16 | 4.364<br>20 | 4.351<br>21 | 5. Implement instruction that is based on designed unit and lesson plans. | 4.436<br>11 | 4.227<br>7 | 4.242<br>8 | 8.7 |

*(T) = tie

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | tem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 17.7 | 4.479<br>19 | 4.476<br>17 | 4.468<br>17 | 6. Construct and use tests and other instruments to monitor student progress. | 4.403<br>13(T)* | 3.882<br>15 | 3.984<br>20 | 19.3 |
| 14.3 | 4.532<br>17 | 4.714<br>15(T)* | 4.725<br>11 | 7. Report student progress to students and parents. | 4.324<br>17 | 3.600<br>22 | 4.172<br>15 | 18 |
| 20.3 | 4.124<br>22 | 4.286<br>21 | 4.429<br>19 | 8. Design and/or incorporate materials, illustrations and examples into lessons that blend with the cultural environments of students. | 4.209<br>20 | 3.833<br>18 | 40.20<br>19 | 19 |
| 3.7 | 4.832<br>4 | 4.954<br>3(T)* | 4.827<br>4 | 9. Use a variety of teaching skills (explains, questions, reinforces, etc.) | 4.565<br>5 | 4.143<br>10 | 4.239<br>9 | 8 |

*(T) = tie

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | Stem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 9.7 | 4.701<br>9 | 4.773<br>10(T)* | 4.733<br>10 | 10. Use a variety of instructional strategies (gaming, simulations, lecture, small group, individualized instruction, e.g.) | 4.393<br>15 | 3.950<br>14 | 4.185<br>12 | 13.7 |
| 3 | 4.878<br>2 | 4.909<br>6(T)* | 4.890<br>1 | 11. Communicate with students. | 4.684<br>2 | 4.454<br>4 | 4.330<br>3 | 3 |
| 15.3 | 4.536<br>15 | 4.714<br>15(T)* | 4.607<br>16 | 12. Communicate with other teachers. | 4.514<br>7 | 4.474<br>3 | 4.183<br>13 | 7.7 |
| 12 | 4.612<br>12 | 4.773<br>10(T)* | 4.611<br>14(T)* | 13. Communicate with supervisors. | 4.438<br>10 | 4.546<br>2 | 4.317<br>4 | 5.3 |
| 11.7 | 4.543<br>14 | 4.762<br>12 | 4.738<br>9 | 14. Communicate with parents. | 4.308<br>18 | 4.000<br>12(T)* | 3.819<br>21 | 17 |
| 21 | 4.310<br>21 | 4.273<br>22 | 4.407<br>20 | 15. Operate and use when appropriate audio-visual aids (e.g., projectors, recorders). | 4.369<br>16 | 3.857<br>17 | 4.253<br>6 | 13 |
| 4.7 | 4.801<br>5 | 5.00<br>1(T)* | 4.751<br>8 | 16. When necessary, changing plans based on the feedback received from students. | 4.478<br>9 | 3.864<br>16 | 4.182<br>14 | 13 |

*(T) = tie

TABLE 5

| Perceived Importance | | | | | Rating of Student Teacher Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Avg. Rating | SBTEs | University Supervision | Student Teacher | tem | Student Teacher | University Supervision | SBTEs | Avg. Rating |
| 11 | 4.757<br>8 | 4.727<br>13(T)* | 4.684<br>12 | 17. Use a variety of classroom management (discipline) strategies. | 4.246<br>19 | 3.737<br>10 | 3.765<br>22 | 20 |
| 9 | 4.695<br>11 | 4.857<br>9 | 4.762<br>7 | 18. React to the needs of students parents, supervisors, other teachers and myself. | 4.403<br>13(T)* | 4.190<br>8 | 4.155<br>16 | 12.3 |
| 4 | 4.842<br>3 | 4.909<br>6(T)* | 4.874<br>3 | 19. Demonstrate openness and flexibility in teaching. | 4.640<br>4 | 4.273<br>6 | 4.265<br>5 | 5 |
| 4.7 | 4.795<br>6 | 4.954<br>3(T)* | 4.825<br>5 | 20. Exhibit a willingness to change when presented with valid data. | 4.652<br>3 | 4.409<br>5 | 4.344<br>2 | 3.3 |
| 1.3 | 4.885<br>1 | 5.000<br>1(T)* | 4.881<br>2 | 21. Demonstrate a willingness to continually improve as a teacher. | 4.787<br>1 | 4.591<br>1 | 4.547<br>1 | 1 |
| 6 | 4.782<br>7 | 4.952<br>5 | 4.801<br>6 | 22. Demonstrate adequate content knowledge. | 4.527<br>6 | 4.000<br>12(T)* | 4.207<br>10 | 9.3 |



*(T) = tie

student teacher performance items were ranked, giving a ranked order statement of student teacher performance. The highest rated SBTE rating of student teachers was on Item 21 (Willingness to Continually Improve as a Teacher). The lowest rating was on Item 2 (Apply Knowledge of Learning in Planning). Thus, for the SBTE ratings, Item 21 would be ranked 1st; Item 2 would be ranked 22nd.

Perceived Importance. While all items were rated quite highly by all groups, the highest three ranked items were:

No. 21. Demonstrate a willingness to continually improve as a teacher

No. 11. Communicates with students

No. 9. Use a variety of teaching skills.

The three lowest ranked items were:

No. 3. Writes clearly stated lesson and unit objectives

No. 8. Design/incorporate materials, etc., that blend with multi-cultural environment of students

No. 15. Operate and use A-V aids.

Student Teacher Performance. There is a much higher variance of raters' opinions of student teacher success on each of the items. However, the three items on which student teachers received highest ratings are:

No. 21. Demonstrate a willingness to continually improve as a teacher

No. 11. Communicate with students

No. 20. Exhibit a willingness to change when presented with valid data.

Five items appear to be about equal with respect to the lowest student ratings. These are:

No. 8. Design/incorporate materials, etc., that blend with multi-cultural environment of students

No. 6. Construct and use tests and other instruments to monitor student progress

No. 1. Apply knowledge of human growth and development in planning for instruction

No. 2. In planning for instruction apply knowledge about how students learn

No. 17. Use a variety of classroom management (discipline) strategies.

In looking at the ranked elements there are some interesting points:

1. In several instances, PTPP students excell at those competencies considered most important by the raters (communication with students, willingness to change).
2. In one case, students appear to perform poorly in the design and incorporation of multicultural elements. This was also a low ranking competency statement.
3. The element ranked lowest in terms of student performance (classroom management) is the element that appears in other success studies.[5] It appears that this is almost a nationwide concern based on published reports and informal sharing of teacher educators.
4. There is no indication in the data where the students feel they gained in competence. It is tempting to indicate that the success is a demonstration of programmatic impact. However, alternative explanations for a number of elements must also be considered. Given the evidence of a high level of interpersonal skills, it may well be that students gain competence in both program and extra-program experiences. Regardless, there is some high degree of success indication provided in the data.

[5]Fred L. Pigge, "Teacher Competencies: Need, Proficiency, and Where Proficiency was Developed," Journal of Teacher Education 29, No. 4, July-August, 1978.

## Evaluation of the Evaluation

Generally, we are pleased with the kinds of data we collected on this first attempt. Our major problem was the long time delay between when the questionnaire data were collected - Spring 1977 - and when they were analyzed - Winter 1977. The reasons for this delay related to the coding of data, key punching, computer programming, and computer operations. The ultimate criterion of our effectiveness is whether the data were used and whether programmatic changes called for by the data actually occurred. On both accounts the answer was "Yes."